The Annals of Statistics
1973, Vol. 1, No. 4, 745-749

ON A THEOREM OF BAHADUR ON THE RATE OF CONVERGENCE OF POINT ESTIMATORS

BY JAMES C. FU

University of Toronto

In this paper, we have proved a fundamental property of the characteristic function for the random variable $(\partial/\partial\theta)\log f(x \mid \theta)$. Based on this result, we have proved under regularity conditions different from Bahadur's that certain classes of consistent estimators $\{\theta_n^*\}$ are asymptotically efficient in Bahadur's sense

$$\lim_{\varepsilon\to 0} \lim_{n\to\infty} \frac{1}{n\varepsilon^2} \log P_\theta\{|\theta_n^* - \theta| \geqq \varepsilon\} = -\frac{I(\theta)}{2}.$$

Our proof also gives a simple and direct method to verify Bahadur's [2] result.

1. Introduction. Let $(\mathscr{X}, \beta, P_\theta)$ be a probability measure space for each θ belonging to the parameter space Θ; here $\mathscr{X}$ is an abstract sample space of point x and Θ is an open interval of the real line. We assume that for each $\theta \in \Theta$, P_θ admits a density function f_θ with respect to a given σ-finite measure. Let $s = (x_1, x_2, \cdots)$ be a sequence of independent, identically distributed (i.i.d.) observations of $\mathscr{X}$. We denote the n-fold Cartesian product space by $(\mathscr{X}^n, \beta^n)$ and the n-fold product probability measure $P_\theta \times P_\theta \times \cdots \times P_\theta$ by $P_\theta^{(n)}$. Thus, s is distributed according to $P_\theta^{(\infty)}$ on $(\mathscr{X}^\infty, \beta^\infty)$. Whenever our intention is clear from the context, we shall leave off the superscript. Thus $P_\theta^{(\infty)}$ will often be written as P_θ.

Let g be a nonnegative function on Θ; g can be interpreted as the density, with respect to Lebesgue measure, of the prior distribution of θ. An estimator (sequence of estimators) θ_n^* is called a maximum probability estimator (MPE) with respect to g if (for each $n = 1, 2, \cdots$) θ_n^* is a β^n-measurable mapping from $\mathscr{X}^n$ to the parameter space Θ and

$$(1.1)\qquad g(\theta_n^*(x_1, \cdots, x_n)) \prod_{i=1}^n f(x_i \mid \theta_n^*(x_1, \cdots, x_n)) = \max_{\theta\in\Theta} [g(\theta) \prod_{i=1}^n f(x_i \mid \theta)],$$

for every $(x_1, \cdots, x_n) \in \mathscr{X}^n$. In the special case when $g(\theta) \equiv 1$, θ_n^* is usually called a maximum likelihood estimator (MLE); in this special case, we use the notation $\hat{\theta}_n$ for the MPE.

Bahadur [1], [2] has shown (under certain regularity conditions) that for any consistent estimator $T_n(s) = T_n(x_1, \cdots, x_n)$,

$$(1.2)\qquad \liminf_{\varepsilon\to 0} \liminf_{n\to\infty} (1/n\varepsilon^2) \log P_\theta\{|T_n(s) - \theta| \geqq \varepsilon\} \geqq -\tfrac{1}{2}I(\theta),$$

and that the maximum likelihood estimator $\hat{\theta}_n$ satisfies

$$(1.3)\qquad \lim_{\varepsilon\to 0} \lim_{n\to\infty} (1/n\varepsilon^2) \log P_\theta\{|\hat{\theta}_n(s) - \theta| \geqq \varepsilon\} = -\tfrac{1}{2}I(\theta),$$

Received September 1971; revised October 1972.

where $I(\theta)$ is Fisher's information. In words: For any consistent estimator T_n, $P_\theta\{|T_n(s) - \theta| \geqq \varepsilon\}$ cannot tend to zero faster than the exponential rate given by $\exp\{-\frac{1}{2}n\varepsilon^2 I(\theta)\}$, and for the maximum likelihood estimator $\hat{\theta}_n$, $P_\theta\{|\hat{\theta}_n(s) - \theta| \geqq \varepsilon\}$ does tend to zero nearly at this optimal exponential rate.

In this paper, we show under simpler, though stronger regularity conditions than Bahadur's that the maximum probability estimator θ_n^* with respect to g is asymptotically efficient in Bahadur's sense almost regardless of the prior density $g(\theta)$. In particular, our main theorem provides a simple method of verifying Bahadur's [1] result (1.3) in certain cases.

2. Conditions and main theorem. Let $l(x \mid \theta) = \log f(x \mid \theta)$ and $l^{(i)}(x \mid \theta) = (\partial/\partial\theta)^i l(x \mid \theta)$ for every $x \in \mathscr{X}$. The following are sufficient conditions for our main theorem.

CONDITION 1. For all x, $l(x \mid \theta)$ is a continuous function of θ. Further, for every $\theta \in \Theta$, there exists a δ-neighborhood of θ, say $N(\theta, \delta)$, and a measurable function $A(x, \theta)$, $E_\theta A^2(x, \theta) < \infty$, such that

$$|l(x \mid \theta') - l(x \mid \theta'')| < A(x, \theta)|\theta'' - \theta'| ,$$

for all $\theta', \theta'' \in N(\theta, \delta)$.

CONDITION 2. For each x, $l(x \mid \theta)$ has a continuous first derivative over Θ. For each $\theta \in \Theta$, $P_\theta\{l^{(1)}(x \mid \theta) \neq 0\} > 0$.

CONDITION 3. For each θ, there exist two constants $u = u(\theta) > 0$ and $v = v(\theta) > 0$ such that $P_\theta\{l^{(1)}(x \mid \theta + \gamma) < 0\} > 0$ and $P_\theta\{l^{(1)}(x \mid \theta + \gamma) > 0\} > 0$, for all γ, $-u < \gamma < u$, and such that the moment generating function $\phi(t, \theta, \gamma) = E_\theta[e^{tl^{(1)}(x \mid \theta+\gamma)}]$ is finite for all (t, γ), $-v < t < v$, $-u < \gamma < u$.

CONDITION 4. The second partial derivative $(\partial/\partial t)^2\phi(t, \theta, \gamma)$ is jointly continuous in t and γ for $-v < t < v$, $-u < \gamma < u$. [Note: The existence of all partial derivatives

$$\left(\frac{\partial}{\partial t}\right)^i \phi(t, \theta, \gamma) , \qquad i = 1, 2, \cdots , \quad \text{for} \quad \phi(t, \theta, \gamma)$$

for $-v < t < v$, $-u < \gamma < u$, follows from Condition 3. This is not enough, however, to show that $(\partial/\partial t)^2\phi(t, \theta, \gamma)$ is *jointly* continuous in t and γ.]

CONDITION 5. The second partial $(\partial/\partial\gamma)(\partial/\partial t)\phi(t, \theta, \gamma)$ exists and is continuous in (t, γ) for $-v < t < v$, $-u < \gamma < u$. Further, $l^{(2)}(x \mid \theta) = (\partial/\partial\theta)^2 l(x \mid \theta)$ exists for all $x \in \mathscr{X}$ and all θ, and

$$\frac{\partial}{\partial\gamma}\frac{\partial}{\partial t}\phi(t, \theta, \gamma) = \int_{\mathscr{X}} \frac{\partial}{\partial\gamma} l^{(1)}(x \mid \theta + \gamma) \exp\{tl^{(1)}(x \mid \theta + \gamma)\}\, dP_\theta(x) ,$$

for all $-v < t < v$, $-u < \gamma < u$.

CONDITION 6. $g(\theta)$ is positive and differentiable over Θ.

CONDITION 7. $\bar{l}_n(s \mid \theta) = l_n(s \mid \theta) + \log g(\theta)$ is concave in θ for every $n \geqq n_0$, n_0 some fixed positive integer, and for all $s = (x_1, x_2, \cdots) \in \mathscr{X}^\infty$.

We also assume that the maximum probability estimator θ_n^* with respect to $g(\theta)$ exists. Sufficient conditions for the existence of θ_n^*, and other properties of θ_n^* can be found in Fu and Gleser [6]. The main theorem of the present paper is as follows.

MAIN THEOREM. *Assume that Conditions* 1 *through* 7 *are satisfied. Then*

$$\lim_{\varepsilon\to 0} \lim_{n\to\infty} \frac{1}{n\varepsilon^2} \log P_\theta\{|\theta_n^*(s) - \theta| \geqq \varepsilon\} = -\tfrac{1}{2}I(\theta) , \tag{2.1}$$

for all θ *in* Θ.

It is worth noting that the theorem is valid for any choice of g for which Conditions 6 and 7 are satisfied. The conclusion (2.1) can be shown to hold also under basically the same regularity conditions on the density $f(x \mid \theta)$ as were assumed by Bahadur ([1], [2]) to prove (2.1) in the special case when $g(\theta) \equiv 1$, plus added assumptions on $g(\theta)$; see Remark 2 at the end of Section 3. However, we take a different approach here.

3. Proof of main theorem. We begin the proof by proving some needed lemmas. The following lemma, due to Daniels [5], gives expansions for $E_\theta l^{(1)}(x \mid \theta')$ and $E_\theta[l^{(1)}(x \mid \theta')]^2$ for θ' in a small neighborhood of θ.

LEMMA 1. *If Condition* 1 *and Condition* 2 *hold, then if* θ' *is sufficiently close to* θ, *the following statements are true*:

(a) $E_\theta l^{(1)}(x \mid) = 0$,

(b) $0 < E_\theta[l^{(1)}(x \mid \theta)]^2 = I(\theta) < \infty$,

(c) $E_\theta l^{(1)}(x \mid \theta') = -(\theta' - \theta)I(\theta) + o(\theta' - \theta)$,

(d) $E_\theta[l^{(1)}(x \mid \theta')]^2 = I(\theta) + o(1)$.

PROOF. See Daniels [5].

Let $\theta' = \theta + \gamma$, where γ is a small (positive or negative) constant. Under Conditions 1 and 2, we have (Lemma 1) $E_\theta l^{(1)}(x \mid \theta + \gamma) = -\gamma I(\theta) + o(\gamma)$, and it is known (Bahadur [3]) that if for small enough γ, $P_\theta\{l^{(1)}(x \mid \theta + \gamma) > 0\} > 0$, then for such γ there exists a unique solution, say $t = \tau_\theta(\gamma)$, of the equation

$$\frac{\partial}{\partial t} \phi(t, \theta, \gamma) = \phi^{(1)}(t, \theta, \gamma) = 0 .$$

The following lemma gives a basic property of $\tau_\theta(\gamma)$.

LEMMA 2. *Assume that Conditions* 1 *through* 5 *are satisfied. Then for each* θ *there exists a unique single-valued function* $\tau_\theta(\gamma)$ *defined on* $-u < \gamma < u$ *such that*

$$\frac{\partial}{\partial t} \phi(t, \theta, \gamma)|_{t=\tau_\theta(\gamma)} \equiv \phi^{(1)}(\tau_\theta(\gamma), \theta, \gamma) = 0 ,$$

and $\tau_\theta(\gamma) = \gamma + o(\gamma)$ *as* $\gamma \to 0$.

PROOF. Let $u(t, \gamma) = \phi^{(1)}(t, \theta, \gamma)$. From Condition 3, it follows (see Parzen [7], page 216) that

$$u(t, \gamma) = \int_{\mathscr{X}} l^{(1)}(x \mid \theta + \gamma) \exp\{t l^{(1)}(x \mid \theta + \gamma)\} \, dP_\theta(x) , \tag{3.1}$$

for $-v < t < v$, $-u < \gamma < u$. By Conditions 1 through 5, Lemma 1, and the fundamental theorem of implicit functions (see, e.g., Taylor [8]) applied to $u(t, \gamma)$, we have $u(\tau_\theta(\gamma), \gamma) = 0$, $\tau_\theta(0) = 0$ and $\tau_\theta'(0) = 1$. The result $\tau_\theta(\gamma) = \gamma + o(\gamma)$ follows immediately from the Mean Value Theorem (expanding $\tau_\theta(\gamma)$ around $\gamma = 0$). This completes the proof.

PROOF OF THEOREM. Let $\rho_1(\theta, \varepsilon, a) = \inf_{t \geqq 0} e^{-at}\phi(t, \theta, \varepsilon)$ and $\bar{l}_n^{(1)}(s \mid \theta) = (d/d\theta)\bar{l}_n(s \mid \theta)$ for every $s \in \mathscr{X}^\infty$. From Condition 2, Condition 6 and Condition 7 we know that

$$\begin{aligned} \bar{l}_n^{(1)}(s \mid \theta') &\geqq 0 , \qquad \text{when} \quad \theta' \leqq \theta_n^*(s) , \\ &\leqq 0 , \qquad \text{when} \quad \theta' \geqq \theta_n^*(s) , \end{aligned}$$

for all $s \in \mathscr{X}^\infty$, $\theta' \in \Theta$. Hence we have

$$P_\theta\{\bar{l}_n^{(1)}(s \mid \theta') > 0\} \leqq P_\theta\{\theta_n^* \geqq \theta'\} \leqq P_\theta\{\bar{l}_n^{(1)}(s \mid \theta') \geqq 0\} .$$

Taking $\theta' = \theta + \varepsilon$ and $a_n = -n^{-1}(\partial/\partial\theta) \log g(\theta)|_{\theta=\theta'}$, $\varepsilon > 0$, we have

$$\begin{aligned} P_\theta \left\{\frac{1}{n} \sum_{i=1}^n l^{(1)}(x_i \mid \theta + \varepsilon) > a_n\right\} &\leqq P_\theta\{\theta_n^* - \theta \geqq \varepsilon\} \\ &\leqq P_\theta \left\{\frac{1}{n} \sum_{i=1}^n l^{(1)}(x_i \mid \theta + \varepsilon) \geqq a_n\right\} . \end{aligned}$$

Since $\lim_{n\to\infty} a_n \to 0$, it follows from the continuity of $\rho_1(\theta, \varepsilon, a)$ at $a = 0$ and from the Bernstein–Chernoff–Bahadur theorem (see Chernoff [4] or Bahadur [3]) that

$$\lim_{n\to\infty} \frac{1}{n} \log P_\theta\{\theta_n^*(s) - \theta \geqq \varepsilon\} = \log \rho_1(\theta, \varepsilon, 0) .$$

In a similar way, we have

$$\lim_{n\to\infty} \frac{1}{n} \log P_\theta\{\theta_n^*(s) - \theta \leqq -\varepsilon\} = \log \rho_2(\theta, \varepsilon, 0) ,$$

where $\rho_2(\theta, \varepsilon, 0) = \inf_{t \leqq 0} \phi(t, \theta, -\varepsilon)$.
It now follows that

$$\lim_{n\to\infty} \frac{1}{n} \log P_\theta\{|\theta_n^* - \theta| \geqq \varepsilon\} = \log \rho(\theta, \varepsilon, 0) , \qquad \text{where} \quad \rho = \max\{\rho_1, \rho_2\} .$$

It will now suffice to show that

$$\lim_{\varepsilon\to\infty} \frac{1}{\varepsilon^2} \log \rho(\theta, \varepsilon, 0) = -\tfrac{1}{2}I(\theta) . \tag{3.2}$$

By Condition 4, Lemma 1 and Lemma 2, we have

$$\begin{aligned} \log \rho_1(\theta, \varepsilon, 0) &= \log \inf_{t \geqq 0} \phi(t, \theta, \varepsilon) = \log \phi(\tau_\theta(\varepsilon), \theta, \varepsilon) \\ &= k_1(\theta, \varepsilon)\tau_\theta(\varepsilon) + \tfrac{1}{2}k_2(\theta, \varepsilon)\tau_\theta^2(\varepsilon) + o(\tau_\theta^2(\varepsilon)) \\ &= -\frac{I(\theta)}{2}\varepsilon^2 + o(\varepsilon^2) \end{aligned} \tag{3.3}$$

where $k_1(\theta, \varepsilon)$ and $k_2(\theta, \varepsilon)$ are the first two cumulants, under P_θ, of the random variable $l^{(1)}(x \mid \theta + \varepsilon)$. It follows from (3.3) that (3.2) holds for $\rho = \rho_1$. It follows from symmetry that (3.2) holds also for $\rho = \rho_2$. It now follows that (3.2) holds for $\rho = \max\{\rho_1, \rho_2\}$. This completes the proof.

REMARK 1. Part of the above proof depends heavily on Condition 7. This condition is rather restrictive, but when $g(\theta)$ is log-concave, it is satisfied for some distributions (such as the normal distribution with known variance, and other members of the Koopmans–Darmois class of probability distributions) of interest in both statistical theory and practice.

REMARK 2. The conclusion can be established also by using Bahadur's [1] method under his regularity conditions, plus some regularity conditions on $g(\theta)$—such as that $(d/d\theta)^i \log g(\theta)$ exists and $\sup_{\theta \in \Theta} |(d/d\theta)^i \log g(\theta)| < \infty$ for $i = 1, 2, 3$. This approach is, however, quite complicated in its details.

4. Acknowledgment. The research reported here was performed as part of the Ph. D. dissertation of the author at the Johns Hopkins University under the direction of Professor L. J. Gleser, whose guidance and many suggestions are gratefully acknowledged. Also the author wishes to thank the referee for many helpful comments.

REFERENCES

[1] BAHADUR, R. R. (1960). Asymptotic efficiency of tests and estimates. *Sankhyā* **22** 229–252.

[2] BAHADUR, R. R. (1967). Rates of convergence of estimates and test statistics. *Ann. Math. Statist.* **38** 303–324.

[3] BAHADUR, R. R. (1971). *Some Limit Theorems in Statistics*. Regional Conference Series in Applied Mathematics, No. 4, SIAM, Philadelphia.

[4] CHERNOFF, H. (1952). A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Ann. Math. Statist.* **23** 493–507.

[5] DANIELS, H. E. (1961). The asymptotic efficiency of a maximum likelihood estimator. *Proc. Fourth Berkeley Symp. Math. Statist. Prob.* **1** 151–164.

[6] FU, JAMES C. and GLESER, LEON JAY (1971). Classical asymptotic properties of a certain estimator related to the maximum likelihood estimator. To appear in *Ann. Inst. Statist. Math.*

[7] PARZEN, E. (1960). *Modern Probability Theory and Its Applications*. Wiley, New York.

[8] TAYLOR, A. E. (1955). *Advanced Calculus*. Blaisdell, Waltham, Mass.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF TORONTO
TORONTO, 181, CANADA